Parsimonious modeling with Information Filtering Networks: construction of predictive graphical models from large numbers of heterogeneous data

Tomaso Aste
&
Tiziana Di Matteo

UCL, Computer Science, Financial Computing & Analytics Group
Use financial networks for predictive modeling

Prediction
Prediction is very difficult, especially about the future. (Niels Bohr)
Predictive modeling

Observation

Model

Parsimony

Prediction
Predictive modeling

Prediction is the estimation of the probability of a (future) event given the available information about other (past) events

\[ p(X_B \mid X_A) \]

We must estimate from data the most likely probability distribution of the system of events

\[ p(X_B \mid X_A) = \frac{p(X_A, X_B)}{p(X_A)} \]

Bayes’ formula

Joint probability

High dimensional problem! (especially for big data)
Prediction is not only about the future,

from

$$p(X_B \mid X_A)$$

we can predict the values of the variables $X_B$ for any kind of scenario of the variables $X_A$

We can estimate the effects of events in $X_A$ on $X_B$
Predictive modeling

The conditional probability

\[ p(X_B \mid X_A) \]

is a tool for:
- test hypothesis
- quantify risk
- stress testing
- analyze scenarios
Predictive modeling

Predicted future values of variables $X_B$ given past values of $X_A$ are the expectation values

$$E[X_B | X_A] = X_B p(X_B | X_A)$$

This is the regression and for linear models (multivariate Gaussian) this is the linear regression formula

Uncertainty about the future given the past is quantified by the conditional entropy

$$H(X_B | X_A) = p(X_B, X_A) \log p(X_B | X_A)$$

$X_A, X_B$
The **reduction of uncertainty** on variables $X_B$ given the knowledge of the past of variables $X^{-}_A$ discounting for their past $X^{-}_B$ is

$$H(X_B | X_B) \quad H(X_B | X_A, X_B) = TE(X_A \rightarrow X_B)$$

This is the **transfer entropy** that for linear models (multivariate Gaussians) coincides with **Granger causality**
Graphical models

To construct the joint multivariate distribution we make use of the structure of conditional dependency

\[
p(X_A, X_B | \tilde{X}) = p(X_A | \tilde{X}) p(X_B | \tilde{X})
\]

\[
p(X_A, X_B | \tilde{X}) \neq p(X_A | \tilde{X}) p(X_B | \tilde{X})
\]

\[
\tilde{X} = X \backslash \{X_A, X_B\}
\]
Graphical models

If these inference networks are chordal (or decomposable) we then have

$$p(X) = \frac{p(X_{\text{cliques}})}{\prod_{\text{cliques}} p(X_{\text{cliques}})^{k_s - 1}}$$

The joint probability distribution of the entire system (large number of variables) can be estimated form the probability distributions of cliques and separators (small number of variables)


This is great… however to establish conditional dependency

\[ p(X_A, X_B \mid \tilde{X}) \neq p(X_A \mid \tilde{X}) p(X_B \mid \tilde{X}) \]

is very hard… actually it is as hard as computing the entiere joint distribution function!

Building the exact inference network is an impossible task for a large number of variables
To solve this problem we propose to build the inference structure for the graphical model as an Information filtering network.


Information filtering networks

Connect the nearest vertices
- Euclidean distance = most correlated
- Hyperbolic distance = mutual information

Keep the graph chordal
- Clique forests

Add other constraints
- Max clique size (2 = MST)
- Planarity (TMFG)
- Information criteria (e.g., Akaike)

These are fast algorithms $O(N^2)$.

(topological & homological measures, betty numbers, cycles and cliques retrieved from construction)

Clique forest

\[ p(X) = \frac{p(X_{\text{cliques}})}{p(X_{\text{separators}})^{k_s - 1}} \]
\[ p(X) = \frac{\prod_{\text{cliques}} p(X_{\text{cliques}})}{\prod_{\text{separators}}^{k_s} p(X_{\text{separators}}}^{k_s} \]
Clique forest

\[ p(X) = \frac{p(X_{\text{cliques}})}{\prod_{k=1}^{s-1} p(X_{\text{separators}})} \]

- \( p(X_{\text{cliques}}) \)
- \( \prod_{k=1}^{s-1} p(X_{\text{separators}}) \)
Graphical Model

Only low-dimensional local probabilities must be estimated.

\[ p(X) = \frac{p(X_{\text{cliques}})}{\bigoplus_{\text{cliques}}} \frac{p(X_{\text{separators}})}{k_s} \]
Prediction

Parsimony

Predictive modeling

\[
p(X) = \frac{p(X_{\text{cliques}})}{\prod_{\text{cliques}} p(X_{\text{separators}})^{k_s}}
\]
By constraining the model to reproduce observed moments while maximizing Shannon-Gibbs entropy (maximum Entropy method), at the second order, we have that the model must be a multivariate Gaussian:

\[ p(X_1, \ldots, X_N) = \frac{1}{Z} \exp \left( \sum_{i,j} X_i J_{i,j} X_j \right) \]

We keep only the significant interactions and set to zero (Max Ent.) the uncertain ones: \( J_{i,j} = 0 \) iff \( X_i, X_j \) conditionally independent. 

\( J_{i,j} \) is sparse and it has the structure given by the information filtering network.
$J_{i,j}$ is computed from local inversion of the covariance matrix over the clique forest

$$J_{i,j} = (C)^{-1} (k_S - 1) (S)^{-1}$$

Cliques $C$ and Separators $S$

We obtain a **sparse inverse covariance** (our graphical model) by doing **local inversion only**

Super-fast algorithm $O(N)$

*even $O(\log N)$ if parallelized*

Test:
is our model, built form past observations, associated with a large likelihood for future observations?

**Statistical description**

\[ p(\text{System}) = \frac{p(\text{cliques})}{p(\text{separators})^{k_s}} \]

**Past**

**Future**

**Observation**

**Model**

**Prediction & Testing**
Test:

is our model, built form past observations, associated with a large likelihood for future observations?
is our model, built from past observations, associated with a large likelihood for future observations?

**Test:**

Linear Model

State-of-the-art sparse model (G-lasso)
Test:

is our model, built from past observations, associated with a large likelihood for future observations?

---

LoGo - TMFG

Linear Model

State-of-the-art sparse model (G-lasso)
Test:
is our model, built form past observations, associated with a large likelihood for future observations?

Good modeling (better than others)

LoGo - TMFG

LoGo - MST

Linear Model

State-of-the-art sparse model (G-lasso)
In which sense we predict?

With \( p(\text{X}_{\text{future}}|\text{X}_{\text{past}}) \) we can predict the future

This is the same as (linear) regression

\[
E[\text{X}_B | \text{X}_A] = \text{X}_B p(\text{X}_B | \text{X}_A) = \text{J}^{-1}_B \text{J}_B \text{X}_A
\]

\text{X}_B

and also Granger causality (2x Transfer entropy)

\[
G(\text{X}_A \rightarrow \text{X}_B) = \log | \text{J} | \log | \text{J}^{-1} | \text{J}
\]

The advantage is that we have a \textit{sparse model} computed in a very efficient way applicable to big-data predictive analytics
Uncertainty spillover across regions in banking system

2005-2015
115 banks:
10 in NA , 66 in EU, 39 in AS
Uncertainty spillover across regions in banking system

2005-2015
Test:
Uncertainty spillover across regions in banking system

\[ I(X_A;X_B) = \frac{1}{2} \log |J_{BB}| - \frac{1}{2} \log |J_{BA} J_{AA} J_{AB}| \]

2005-2015
Test:

Uncertainty spillover across regions in banking system

2005-2015

Aggregation of risk is straightforward

\[
I(X_A;X_B) = \frac{1}{2} \log |J_{BB}| + \frac{1}{2} \log |J_{BA}| + \frac{1}{2} \log |J_{AB}|
\]

\[
TE(X_A \rightarrow X_B) = \frac{1}{2} \log |J_{BB}| - \frac{1}{2} \log |J_{BB} - J_{BA} - J_{AB}|
\]

\[
I(X_A;X_B) = \frac{1}{2} \log |J_{BB}| - \frac{1}{2} \log |J_{BB} - J_{BA} - J_{AB}|
\]

\[
TE(X_A \rightarrow X_B) = \frac{1}{2} \log |J_{BB}| - \frac{1}{2} \log |J_{BB} - J_{BA} - J_{AB}|
\]

TE = 0.016

TE = 0.066
Test:
Uncertainty spillover across regions in banking system

2005-2008

$\text{TE} = 0.014$

$\text{TE} = 0.150$

$\text{TE} = 0.048$

$I(X_A;X_B) = \frac{1}{2} \log |J_{BB}| - \frac{1}{2} \log |J_{BA} J_{AA} J_{AB}|$

$\text{TE}(X_A \rightarrow X_B) = \frac{1}{2} \log |J_{BB}| - \frac{1}{2} \log |J_{BB} J_{BA} J_{AA} J_{AB}|$
Test:
Uncertainty spillover across regions in banking system

\[
I(X_A;X_B) = \frac{1}{2} \log |J_{BB}| - \frac{1}{2} \log |J_{BB} J_{BA} J_{AB}^{-1} J_{AB}^-1|
\]

\[
TE(X_A \rightarrow X_B) = \frac{1}{2} \log |J_{BB}^{-1} J_{BA} J_{AA} J_{AB}^-1|
\]

2008-2012

EU

NA

AS

TE = 0.0

16%

42%

TE = 0.0

TE = 0.0

TE = 0.0
Test:
Uncertainty spillover across regions in banking system

\[ I(X_A;X_B) = \frac{1}{2} \log |J_{BB}| - \frac{1}{2} \log |J_{BA}| - \frac{1}{2} \log |J_{AB}| \]

\[ TE(X_A \rightarrow X_B) = \frac{1}{2} \log |J_{BB}| - \frac{1}{2} \log |J_{BA}| - \frac{1}{2} \log |J_{AB}| \]

2012-2015
With $p(X_B|X_A)$ we can quantify probability of future events.

With $p(X_B|X_A)$ we can predict impact of unobserved scenarios and test hypothesis.

$p(X_A,X_B)$ can be constructed from local probability estimations over an information filtering network (low dimension problem).

Nodes and edges can be added or removed with local moves only.

Aggregation of risk is straightforward.

LoGo works better than state-of-the-art sparse graphical models and it is faster.
W. Barfuss, GP Massara, T Di Matteo & TA


http://www.cs.ucl.ac.uk/staff/tomaso_aste/
http://fincomp.cs.ucl.ac.uk/
http://blockchain.cs.ucl.ac.uk/

Si l’ordre satisfait la raison, le désordre fait les délices de l’imagination
Paul Claudel